Ergemen, Y.E., Velasco, C. (2017). Estimation of fractionally integrated panels with fixed effects and cross-section dependence. Journal of Econometrics, v. 196, n. 2, pp. 248-258.
Available in https://doi.org/10.1016/j.jeconom.2016.05.020.

# Estimation of fractionally integrated panels with fixed effects and cross-section dependence

Yunus Emre Ergemen [a], Carlos Velasco [b,*]

[a] CREATES and Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, Building 2628, 8210 Aarhus V, Denmark
[b] Department of Economics, Universidad Carlos III de Madrid, Calle Madrid, 126, 28903 Getafe (Madrid), Spain

A B S T R A C T

We consider a large $N, T$ heterogeneous panel data model with fixed effects, common factors allowing for cross-section dependence, and persistent data and errors, which are assumed fractionally integrated. We propose individual and common-correlation estimates for the slope parameters while error memory parameters are estimated from regression residuals. The individual parameter estimates are all $\sqrt{T}$ consistent, asymptotically normal and mutually uncorrelated, irrespective of cointegration between defactored observables. A study of small-sample performance and an empirical application to realized volatility persistence are included.

*JEL classification:*
C22
C23

*Keywords:*
Fractional cointegration
Factor models
Long memory
Realized volatility

## 1. Introduction

In macroeconomics and finance, variables are generally presented in the form of panels describing dynamic characteristics of different units such as countries or assets. Some of these macroeconomic panels include GDP, interest, inflation and unemployment rates while in finance, it is standard to use a panel data approach in portfolio performance evaluations. Panel data analyses lead to more robust inference under correct specification since they allow for cross sections to be interacting with each other while also accounting for individual cross-section characteristics. Recent research in panel data theory has mainly focused on dealing with unobserved fixed effects and cross-section dependence in stationary weakly dependent panels. For instance, Pesaran (2006) and Moon and Weidner (2015) propose estimation of general panel data models in which all variables are $I(0)$. The research on nonstationary panel data theory is also abundant. However, those papers that both contain nonstationarity and allow for fixed effects and cross-section dependence are limited to the unit-root case. For example, Kapetanios et al. (2011) extend the study by Pesaran (2006) to panels where observables and factors are integrated $I(1)$ processes while regression errors are $I(0)$. Furthermore, Bai and Ng (2004) and Bai (2010) propose unit-root testing procedures when idiosyncratic shocks and the common factor are both $I(1)$. Similarly, Moon and Perron (2004) propose the use of dynamic factors for unit-root testing for panels with cross-section dependence.

In the same way that many economic time series, such as aggregate output, real exchange rates, equity volatility, asset and stock market realized volatility, have been theoretically and empirically shown to exhibit long-range dependence of non-integer orders, see e.g. Robinson (1978), Granger (1980), Baillie (1996), Gil-Alaña and Robinson (1997) and Bollerslev et al. (2013), panel data models should also be able to accommodate such behavior. However, the study of panel data models with fractional integration characteristics has been completely neglected until very recently, and only a few papers study fractional panels. Hassler et al. (2011) propose a test for the memory parameter under a fractionally integrated panel setup with multiple time series. Robinson and Velasco (2015) propose several estimation techniques for a type-II (i.e. time-truncated) fractionally integrated panel data model with fixed effects.

In this paper, we consider a panel data model that allows for fractionally integrated long-range dependence in both idiosyncratic shocks and a set of common factors. In these models, persistence is described by a memory or fractional integration

* Corresponding author.
*E-mail addresses:* ergemen@econ.au.dk (Y.E. Ergemen), carlos.velasco@uc3m.es (C. Velasco).

parameter, constituting an alternative to dynamic autoregressive (AR) panel data models. This fractional class of modeling nests the standard $I(0)$ and $I(1)$ cases smoothly and removes the necessity for preliminary unit-root or stationarity testing, which may be required in autoregressive modeling. Furthermore, parameter estimates and related test statistics have standard distributions, unlike in the $I(1)$ autoregressive case.

The setup we consider requires that both the number of cross section units, $N$, and the length of the time series, $T$, grow in the asymptotics, departing from the case of multivariate time series (with $N$ fixed) or short panels (with $T$ fixed). Our setup differs from Hassler et al. (2011) and Robinson and Velasco (2015) in that (a) we model cross-section dependence employing an unobservable common factor structure that can be serially correlated and display long-range dependence, which makes the model more general by introducing cross-section dependence without further structural impositions on the idiosyncratic shocks for which identical distribution restriction across cross-section units is also relaxed; (b) our model including covariates allows for, but does not require, fractional cointegration identifying long-run relationships between the unobservable components of the observed time series.

Using a type-II fractionally integrated panel data model with fixed effects and cross-section dependence modeled through a common factor structure, we allow for long-range persistence through the factors and the integrated idiosyncratic shock. This model can be seen as an extended version of the setup of Robinson and Hidalgo (1997) and Robinson and Hualde (2003) to panel data models and of Pesaran (2006) to nonstationary systems with possible cointegration among idiosyncratic components of observed variables, where endogeneity of covariates is driven by the common factor structure independent of those idiosyncratic components. However observed time series can display the same memory level due to dependence on a persistent common factor thereby leading to spurious regressions, but the idiosyncratic component of the error term in the regression equation could be less integrated than the idiosyncratic terms of the covariates, leading to an unobservable cointegrating relationship which can only be disclosed by previously projecting out the multifactor structure.

Our method avoids the spurious regression effect as well as the nonstandard asymptotics and slower rate of convergence that can appear in panels with fractionally integrated time series when the sum of the integration parameters of the regression error and of the regressor is above 1/2, even without common components, by applying enough prewhitening and making the regressors approximately strictly exogenous after projection on cross-section averages.

In the estimation, we first remove the fixed effects, in a standard way, by taking first differences. Then, we use a CSS criterion to estimate possibly heterogeneous slope and memory parameters, where individual time series are projected on (fractionally) differenced cross-section averages of the dependent variable and regressors, leading to GLS type of estimates for the slope parameter. We show that both individual slope and fractional integration parameter estimates are $\sqrt{T}$ consistent, and asymptotically normally distributed. The slope estimates have an asymptotic Gaussian distribution irrespective of cointegration among idiosyncratic components of the observables, which are assumed independent of the regression errors, though observables are not.

We explore the performance of our estimation method via Monte Carlo experiments, which indicate that our estimation method has good small-sample properties. Last but not least, we present an application on industry-level realized volatilities using the general model. We analyze how each industry realized volatility is related to a composite market realized volatility measure. We identify several cointegrating relationships between industry and market realized volatilities, which may have direct implications for policy and investment decisions.

Next section details the model and necessary assumptions, explains the estimation strategy, and discusses the asymptotic behavior of the estimates. Section 3 presents Monte Carlo studies for both models. Section 4 contains an application on the systematic macroeconomic risk, employing industry-level realized volatility analysis. Finally, Section 5 concludes the paper. Throughout the paper, we use the notation $(N, T)_j$ to denote joint cross-section and time-series asymptotics, $\to_p$ to denote convergence in probability and $\to_d$ to denote convergence in distribution. All mathematical proofs and technical lemmas are collected in a supplementary online appendix (see Appendix A).

## 2. A fractionally integrated heterogeneous panel data model

In this section we present a heterogeneous panel data model with fixed effects and cross-section dependence where both regression errors and common factors are fractionally integrated and covariates are allowed to be endogenous through those unobserved common factors. Depending on the relationship of the integration orders of the idiosyncratic components of dependent and independent variables and of the common factors, it is possible to describe situations in which there is cointegration between observables or only among the idiosyncratic components of observables once common factors are removed.

For $i = 1, \ldots, N$ and $t = 0, 1, \ldots, T$, the model that generates the observed series $y_{it}$ and $X_{it}$ is given by

$$\begin{aligned} y_{it} &= \alpha_i + \beta_{i0}' X_{it} + \gamma_i' f_t + \lambda_t^{-1}(L; \theta_{i0}) \varepsilon_{it}, \\ X_{it} &= \mu_i + \Gamma_i' f_t + e_{it} \end{aligned} \tag{1}$$

where $X_{it}$ is $k \times 1$, the unobserved common factor $f_t$ is $m \times 1$ with $k$, $m$ fixed, $\gamma_i$, $\Gamma_i$ are vectors of factor loadings, $\varepsilon_{it}$ are idiosyncratic shocks; $\theta_{i0} \in \Theta_i \subset \mathbb{R}^{p+1}$ is a $(p+1) \times 1$ parameter vector; $L$ is the lag operator and for any $\theta \in \Theta$ and for each $t \geq 0$,

$$\lambda_t(L; \theta) = \sum_{j=0}^{t} \lambda_j(\theta) L^j \tag{2}$$

truncates $\lambda(L; \theta) = \lambda_\infty(L; \theta)$. We assume that $\lambda(L; \theta)$ has this particular structure,

$$\lambda(L; \theta) = \Delta^\delta \psi(L; \xi),$$

where $\delta \geq 0$ is a scalar, $\xi$ is a $p \times 1$ vector, $\theta = (\delta, \xi')'$. Here $\Delta = 1 - L$, so that the fractional filter $\Delta^\delta$ has the expansion

$$\Delta^\delta = \sum_{j=0}^{\infty} \pi_j(\delta) L^j, \quad \pi_j(\delta) = \frac{\Gamma(j-\delta)}{\Gamma(j+1)\Gamma(-\delta)},$$

and denote the truncated version as $\Delta_t^\delta = \sum_{j=0}^{t-1} \pi_j(\delta) L^j$, with $\Gamma(-\delta) = (-1)^\delta \infty$ for $\delta = 0, 1, \ldots,$ $\Gamma(0)/\Gamma(0) = 1$; $\psi(L; \xi)$ is a known function such that for complex-valued $x$, $|\psi(x; \xi)| \neq 0$, $|x| \leq 1$ and in the expansion

$$\psi(L; \xi) = \sum_{j=0}^{\infty} \psi_j(\xi) L^j,$$

the coefficients $\psi_j(\xi)$ satisfy the conditions

$$\psi_0(\xi) = 1, \; \left|\psi_j(\xi)\right| = O\left(\exp\left(-c\left(\xi\right) j\right)\right), \tag{3}$$

where $c(\xi)$ is a positive-valued function of $\xi$. Note that

$$\lambda_j(\theta) = \sum_{k=0}^{j} \pi_{j-k}(\delta) \psi_k(\xi), \; j \geq 0, \tag{4}$$

behaves asymptotically as $\pi_j(\delta)$,

$$\lambda_j(\theta) = \psi(1;\xi)\pi_j(\delta) + O(j^{-\delta-2}), \quad \text{as } j \to \infty,$$

see Robinson and Velasco (2015), where

$$\pi_j(\delta) = \frac{1}{\Gamma(-\delta)} j^{-\delta-1}(1 + O(j^{-1})) \quad \text{as } j \to \infty,$$

so the value of $\delta_0$ determines the asymptotic stationarity ($\delta_0 < 1/2$) or nonstationarity ($\delta_0 \geq 1/2$) of $y_{it} - \alpha_i - \beta_{i0}' X_{it} - \gamma_i f_t$ while $\psi(L;\xi)$ describes short memory dynamics as those produced by stable ARMA models.

The variates $\alpha_i$ and $\mu_i$ are covariate-specific fixed effects for which we do not make any particular assumption, and $f_t$ and $e_{it}$ are fractionally integrated of orders $\varrho$ and $\vartheta_i$, respectively, denoted as $f_t \sim I(\varrho)$ and $e_{it} \sim I(\vartheta_i)$, where $\varrho$ and $\vartheta_i$ are nuisance parameters, while the constant parameters $\theta_{i0}$ and $\beta_{i0}$ are the objects of interest, together with mean effects when the slope parameters are assumed to be generated by a random-coefficients model.

When $\psi(L;\xi_{i0}) = 1 - \xi_{i0}L$ corresponds to a finite AR(1) polynomial, the model can be reorganized in terms of the variable $\Delta_t^{\delta_{i0}} y_{it}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ as

$$\begin{aligned}\Delta_t^{\delta_{i0}} y_{it} &= (1 - \xi_{i0})\Delta_t^{\delta_{i0}}\alpha_i + \xi_{i0}\Delta_t^{\delta_{i0}} y_{it-1} + \beta_{i0}'(1 - \xi_{i0}L)\Delta_t^{\delta_{i0}} X_{it} \\ &\quad + \gamma_i(1 - \xi_{i0}L)\Delta_t^{\delta_{i0}} f_t + \varepsilon_{it},\end{aligned}$$

which is then easily comparable to a standard dynamic AR(1) panel data model with cross-section dependence and covariates

$$y_{it} = (1 - \rho_i)\alpha_i + \rho_i y_{it-1} + \beta_{i0}' X_{it} + \gamma_i f_t + \varepsilon_{it}.$$

In both models, error terms are *iid*, and there are fixed effects (so long as $\delta_{i0} \neq 1$, $\xi_{i0} \neq 1$ and $\rho_i \neq 1$), though $\Delta_t^{\delta_{i0}}\alpha_i$ is getting smaller as $t \to \infty$. However, autoregressive panel data models can only cover a limited range of persistence levels, just $I(0)$ or $I(1)$ series depending on whether $|\rho_i| < 1$ or $\rho_i = 1$. On the other hand, the fractional model in (1) covers a wide range of persistence levels depending on the values of $\delta_{i0}$ and $\varrho$, including the unit root case and beyond. In addition, (1) accounts for persistence in cross-section dependence depending on the degree of integration of $\Delta_t^{\delta_{i0}} f_t$.

In the factor models of Pesaran (2006) and Bai (2009), (possibly endogenous) covariates are $I(0)$ so they can only address cases in which there is no long-range dependence in the panel. Kapetanios et al. (2011) study a model where factors and regressors are $I(1)$ processes while errors are stationary $I(0)$ series. Our approach, on the other hand, is specifically geared towards general nonstationary behavior in panels and addresses estimation of both cointegrating and non-cointegrating relationships. We do not explicitly include the presence of observable common factors and time trends in the equations for $y_{it}$ and $X_{it}$, but these could be incorporated and treated easily by our estimation methods as we later discuss.

We introduce the following regularity conditions for the model in (1).

### Assumption A

**A.1.** The idiosyncratic shocks, $\varepsilon_{it}$, $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$ are independently distributed across $i$ and identically and independently distributed across $t$ with zero mean and variance $\sigma_i^2$, and have a finite fourth-order moment, and $\delta_{i0} \in [0, 3/2)$.

**A.2.** The common factor satisfies $f_t = \Delta_t^{-\varrho} z_t^f$, $\varrho < 3/2$, where $z_t^f = \Phi^f(L) v_{t-k}^f$ with $\Phi^f(s) = \sum_{k=0}^{\infty} \Phi_k^f s^k$, $\sum_{k=0}^{\infty} k \left\| \Phi_k^f \right\| < \infty$, $\det\left(\Phi^f(s)\right) \neq 0$ for $|s| \leq 1$ and $v_t^f \sim iid(0, \boldsymbol{\Omega}_f)$, $\boldsymbol{\Omega}_f > 0$, $E\left\| v_t^f \right\|^4 < \infty$, and the idiosyncratic shocks $e_{it}$ are independent in $i$ and satisfy $e_{it} = \Delta_t^{-\vartheta_i} z_{it}^e$, $\sup_i \vartheta_i < 3/2$, where $z_{it}^e = \Phi_i^e(L) v_{it-k}^e$ with $\Phi_i^e(s) = \sum_{k=0}^{\infty} \Phi_{ik}^e s^k$, $\sup_i \sum_{k=0}^{\infty} k \left\| \Phi_{ik}^e \right\| < \infty$, $\det\left(\Phi_i^e(s)\right) \neq 0$ for $|s| \leq 1$ and $v_{it}^e \sim iid(0, \boldsymbol{\Omega}_{ie})$, $\boldsymbol{\Omega}_{ie} > 0$, $\sup_{i,t} E\left\| v_{it}^e \right\|^4 < \infty$.

**A.3.** The covariate-specific idiosyncratic shocks, $e_{it}$, the idiosyncratic error terms, $\varepsilon_{it}$, and the unobservable common factor, $f_t$, are all pairwise independent and independent of $\gamma_i$ and $\Gamma_i$, which are also independent in $i$.

**A.4**. $\operatorname{Rank}(\overline{\mathbf{C}}_N) = m \leq k + 1$ for all $N$, where

$$\overline{\mathbf{C}}_N = \begin{pmatrix} \overline{\boldsymbol{\beta}_0' \boldsymbol{\Gamma}'}_N + \overline{\boldsymbol{\gamma}'}_N \\ \overline{\boldsymbol{\Gamma}'}_N \end{pmatrix}$$

with $\overline{\boldsymbol{\gamma}}_N = N^{-1}\sum_{i=1}^N \gamma_i$, $\overline{\boldsymbol{\Gamma}}_N = N^{-1}\sum_{i=1}^N \Gamma_i$, $\overline{\boldsymbol{\beta}_0' \boldsymbol{\Gamma}'}_N = N^{-1}\sum_{i=1}^N \beta_{i0}' \Gamma_i'$.

**A.5.** For $\xi \in \Xi$, $\psi(x;\xi)$ is differentiable in $\xi$ and, for all $\xi \neq \xi_0$, $|\psi(x;\xi)| \neq |\psi(x;\xi_0)|$ on a subset of $\{x : |x| = 1\}$ of positive Lebesgue measure, and (3) holds for all $\xi \in \Xi$ with $c(\xi)$ satisfying

$$\inf_{\Xi} c(\xi) = c^* > 0. \tag{5}$$

Assumption A.1 is quite standard in panel data literature and relaxes the identical distribution condition across $i$, in particular allowing for each equation error to have different persistence and variance. The condition on the memory parameter $\delta_{i0}$ is motivated by the use of first differences in the methodology. Assumption A.2 states that the factor series and the regressor idiosyncratic terms are multivariate integrated nonsingular linear processes of orders $\varrho$ and $\vartheta_i$, respectively, where the $I(0)$ innovations of $f_t$ are not collinear. We assume that all components of these vectors are of the same integration order to simplify conditions and presentation, though some heterogeneity could be allowed at the cost of making notation much more complex.

Assumption A.3 is a standard condition and does not restrict covariates to be exogenous, because as long as $\Gamma_i \neq 0$ and $\gamma_i \neq 0$, endogeneity will be present. Furthermore, this could be relaxed by assuming $E(X \otimes \varepsilon) = 0$ and finite higher order moments, but this would require more involved derivations and no further insights.

Assumption A.4 introduces a rank condition that simplifies derivations and requires that $k + 1 \geq m$. It is possible that some of our results hold if this condition is dropped, but at the cost of introducing more technical assumptions and derivations, see e.g. Pesaran (2006) and Kapetanios et al. (2011). Furthermore, as pointed out by Pesaran (2006), consistency of individual heterogeneous slope parameters fails when the rank condition does not hold so it is also imposed therein for asymptotic analysis although for the estimation of mean effects, it is not required. This condition facilitates the identification of the $m$ factors using the $k + 1$ cross section averages of the observables and still allows for cointegration among idiosyncratic elements of each unit.

Assumption A.5 ensures that $\psi(L;\xi)$ is smooth for $\xi \in \Xi$, and the weights $\psi_j$ lead to short-memory dynamics as is also assumed by Robinson and Velasco (2015), where the parameter space $\Xi$ can depend on stationarity and invertibility restrictions on $\psi(L;\xi)$.

Under the given set of assumptions, we perform the estimation in first differences to remove fixed effects. Along this line, we caution that attempting to take fractional differences directly would not remove exactly fixed effects but would instead introduce fractional trends because of the truncation of the fractional filter that may lead to further complications in their treatment. For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, the first-differenced model, including only asymptotically stationary variables, is

$$\begin{aligned}\Delta y_{it} &= \beta_{i0}' \Delta X_{it} + \gamma_i' \Delta f_t + \Delta \lambda_t^{-1}(L;\theta_{i0})\varepsilon_{it}, \\ \Delta X_{it} &= \Gamma_i' \Delta f_t + \Delta e_{it}.\end{aligned} \tag{6}$$

The estimation method we propose for each $\beta_{i0}$ is in essence a GLS estimation after prewhitening by means of fractional $\delta^*$ differencing, where $\delta^*$ is a sufficiently large differencing parameter chosen by the econometrician that could be a noninteger. To see this, write

$$\begin{aligned}\Delta_{t-1}^{\delta^*-1}\Delta y_{it} &= \beta_{i0}'\Delta_{t-1}^{\delta^*-1}\Delta X_{it} + \gamma_i'\Delta_{t-1}^{\delta^*-1}\Delta f_t \\ &\quad + \Delta_{t-1}^{\delta^*-1}\Delta\lambda_t^{-1}(L;\theta_{i0})\varepsilon_{it},\end{aligned}$$

the idiosyncratic error term is approximately $\Delta_{t-1}^{\delta^*-\delta_{i0}}\psi(L;\xi_{i0})\varepsilon_{it} \approx I(0)$ when $\delta^* \approx \delta_{i0}$. This prewhitening step extends Bai and Ng (2004)'s method based on first differencing allowing for general levels of persistence in the model without need of exact knowledge on them by the econometrician.

Adapting Pesaran (2006)'s method, we remove the factor structure by projecting the transformed model on the fractionally differenced cross-section averages in order to match the corresponding persistence level. The general intuition is that to control strong persistence, enough differencing is needed in absence of knowledge on the true value of $\delta_{i0}$, e.g. setting $\delta^* = 1$ and working with first differences as done by Bai and Ng (2004). This policy requires that all variables in (6) are (asymptotically) stationary and bears the implicit assumption that variables are not much more persistent than a unit root, while allowing $\delta_{i0}$ to be smaller than $\vartheta_i$, implying a cointegrating relationship between the defactored version of $y_{it}$, $\beta_{i0}'e_{it} + \lambda_t^{-1}(L;\theta_{i0})\varepsilon_{it} \sim I(\max\{\delta_{i0},\vartheta_i\})$, and that of $X_{it}$, $e_{it} \sim I(\vartheta_i)$, when $\vartheta_i > \delta_{i0}$. In case of the presence of incidental linear trends, it would be possible to work with second differences of data, which would remove exactly them at the cost of introducing slightly modified initial conditions for the fractional differences of observed data.

Denote $\mathbf{y}_i = (y_{i1},\dots,y_{iT})'$, $\mathbf{X}_i = (X_{i1},\dots,X_{iT})'$, $\mathbf{F} = (f_1,\dots,f_T)'$, $\mathbf{E}_i = (e_{i1},\dots,e_{iT})'$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1},\dots,\varepsilon_{iT})'$. We can write down the model in first differences as

$$\begin{aligned}\Delta\mathbf{y}_i &= \Delta\mathbf{X}_i\boldsymbol{\beta}_{i0} + \Delta\mathbf{F}\gamma_i + \Delta\lambda_t^{-1}(L;\theta_{i0})\boldsymbol{\varepsilon}_i \\ \Delta\mathbf{X}_i &= \Delta\mathbf{F}\Gamma_i + \Delta\mathbf{E}_i.\end{aligned}$$

Then, the $T\times T$ projection matrix $\bar{\mathcal{W}}_T$ is defined by

$$\begin{aligned}\bar{\mathcal{W}}_T &= \bar{\mathcal{W}}_T(\delta^*) = \mathbf{I}_T - \bar{\mathbf{H}}(\delta^*)(\bar{\mathbf{H}}(\delta^*)'\bar{\mathbf{H}}(\delta^*))^-\bar{\mathbf{H}}(\delta^*)' \\ \bar{\mathbf{H}} &= \bar{\mathbf{H}}(\delta^*) = \frac{1}{N}\sum_{j=1}^N\left(\mathcal{Y}_j(\delta^*)\quad \mathcal{X}_j(\delta^*)\right)\end{aligned}$$

where $(\cdot)^-$ denotes generalized inverse, and $\bar{\mathbf{H}}(\delta^*)$ is the $T\times(k+1)$ matrix of fractionally differenced cross-section averages with

$$\mathcal{Y}_j = \mathcal{Y}_j(\delta^*) = \Delta_{t-1}^{\delta^*-1}\Delta\mathbf{y}_j \quad\text{and}\quad \mathcal{X}_j = \mathcal{X}_j(\delta^*) = \Delta_{t-1}^{\delta^*-1}\Delta\mathbf{X}_j,$$

where we drop $\delta^*$ in the notation when there is no confusion.

Denote $\mathcal{F} = \mathcal{F}(\delta^*) = \Delta_{t-1}^{\delta^*-1}\Delta\mathbf{F}$ and introduce the infeasible projection matrix on unobserved factors

$$\mathcal{W}_f = I_T - \mathcal{F}(\mathcal{F}'\mathcal{F})^-\mathcal{F}'.$$

Adapting Pesaran (2006)'s argument, under the rank conditions in Assumptions A.2 and A.4, as $(N,T)_j \to \infty$, it is possible to approximately eliminate the factors by projecting on $\bar{\mathcal{W}}_T$, based on cross section averages, instead of on $\mathcal{W}_f$, based on unobserved factors, which produces exactly $\mathcal{W}_f\mathcal{F} = 0$, see Pesaran (2006) and the supplementary online material in Appendix A for details on the specific errors introduced by this replacement. That is, both projections can be used interchangeably for factor removal in the asymptotics as long as the rank condition holds. Along this line, the possibility of including observed factors in the covariates as in Pesaran (2006) should also be noted just by enlarging $\bar{\mathbf{H}}(\delta^*)$ with an appropriately fractionally differenced version of such factors. Introducing such observed factors would not alter any of the results since they would also be entirely removed by projection, and, similarly a constant could be added to project out the contribution of the differences of individual linear trends.

The (preliminary) estimate of $\boldsymbol{\beta}_{i0}$ for some fixed $\delta^*$ is given by

$$\hat{\boldsymbol{\beta}}_i(\delta^*) := \left(\mathcal{X}_i'\bar{\mathcal{W}}_T\mathcal{X}_i\right)^{-1}\mathcal{X}_i'\bar{\mathcal{W}}_T\mathcal{Y}_i,$$

where the following identification condition is satisfied.

**Assumption A.6.** $\mathcal{X}_i'\bar{\mathcal{W}}_T\mathcal{X}_i$ and $\mathcal{X}_i'\mathcal{W}_f\mathcal{X}_i$ are full rank for all $i = 1,\dots,N$.

Note that choosing $\delta^* \geq 1$, so that $\vartheta_i + \delta_{i0} - 2\delta^* < 1$ for all possible values of $\vartheta_i$ and $\delta_{i0}$ in the allowed sets prescribed in Assumptions A.1 and A.2, guarantees that all detrended variables are asymptotically stationary and that sample moments converge to population limits as $(N,T)_j \to \infty$. This, together with the identifying conditions in Assumption A lead to the consistency of $\hat{\boldsymbol{\beta}}_i(\delta^*)$, as we show in the next theorem. This does not require further restrictions on the rate at which $N$ and $T$ diverge, just that $\delta^*$ is not smaller than one. This approach is similar to the choice of working with first differences in Bai and Ng (2004) when trying to estimate the common factors from $I(1)$ nonstationary data by principal components although using $\delta^*$ provides greater flexibility in terms of persistence that could be allowed in the model, extending their method based on first differencing.

**Theorem 1.** *Under Assumption* A, $\delta^* \geq 1$, *as* $(N,T)_j \to \infty$, *for fixed i,*

$$\hat{\boldsymbol{\beta}}_i(\delta^*) \to_p \boldsymbol{\beta}_{i0}.$$

This theorem shows the unit-wise convergence result for the slope parameters, which is standard for the study of heterogeneous panels in that the interest is in obtaining unit-specific inference. We later provide results for the estimation of mean effects under a random coefficients model in Section 2.2, where efficiency improvements could be obtained under homogeneity restrictions, subject to control of a second-order bias. The random coefficient model could also be useful for conducting tests of homogeneity.

We next analyze the asymptotic distribution of $\hat{\boldsymbol{\beta}}_i(\delta^*)$ when $\delta^*$ is large enough so that aggregate memory of the idiosyncratic regression error term and regressor component is as small as desired. Define for $\delta^* \geq 1$,

$$\boldsymbol{\Sigma}_{ie}(j) = \sum_{k=0}^\infty \boldsymbol{\Phi}_{ik}^e(\delta^*-\vartheta_i)\boldsymbol{\Omega}_{ie}\boldsymbol{\Phi}_{ij+k}^e(\delta^*-\vartheta_i)',\ \ j=0,1,\dots,$$

$\boldsymbol{\Sigma}_{ie}(j) = \boldsymbol{\Sigma}_{ie}(-j)'$, $j < 0$, the autocovariance function of the detrended idiosyncratic component of the regressors $x_{it}$, where the weights $\boldsymbol{\Phi}_i^e(\delta^*-\vartheta_i) = \sum_{j=0}^k \boldsymbol{\Phi}_{ik-j}^e\pi_j(\delta^*-\vartheta_i)$ incorporate the prewhitening effect, and for $\vartheta_i + \delta_{i0} - 1/2 < 2\delta^*$ (which can be guaranteed by taking $\delta^* > 5/4$) define

$$\boldsymbol{\Sigma}_{i0} = \sum_{j=-\infty}^\infty \boldsymbol{\Sigma}_{ie}(j)\zeta_{i0}(j),$$

where $\zeta_{i0}(j) = \sum_{k=0}^\infty \lambda_k^{-1}(\delta_{i0}-\delta^*,\xi_{i0})\lambda_{k+|j|}^{-1}(\delta_{i0}-\delta^*,\xi_{i0})$, $j = 0,\pm1,\dots,$ is the autocovariance sequence of the prewhitened regression errors, noting that the $\lambda_k^{-1}(\delta_{i0}-\delta^*,\xi_{i0})$ filter is square summable.

Setting $\delta^* = 1$ could be enough to obtain the asymptotic normality of the OLS estimates of $\beta_{i0}$ if we further restrict the aggregate memory as in the next condition. Set

$$\vartheta_{max} = \max_i \vartheta_i,\quad \delta_{max} = \max_i \delta_{i0}.$$

**Assumption B.** $\delta^* \geq 1$ is chosen so that $\max\{\varrho + \delta_{max}, \varrho + \vartheta_{max}, \delta_{max} + \vartheta_{max}, 2\vartheta_{max}\} - 1/2 < 2\delta^*$.

This condition requires that $\delta^*$ is large enough compared to the maximal orders of integration of the idiosyncratic components and factors. Noting that these are assumed to be less than 3/2, it is sufficient to take $\delta^* > 5/4$, while $\delta^* = 1$ is sufficient if the maximal aggregate memory $\max\{\varrho + \delta_{max}, \varrho + \vartheta_{max}, \delta_{max} + \vartheta_{max}, 2\vartheta_{max}\} < 5/2$, which is stronger than the 3 upper bound implied by all memory parameters being less than 3/2, nevertheless allowing for some trade off between them. This choice of $\delta^*$ can be seen as a time-domain prewhitening alternative to the frequency domain GLS procedure of Robinson and Hidalgo (1997). Considering the fact that most indicators in economics and finance satisfy $\delta_{i0} < 3/2$, taking e.g. $\delta^* = 2$ enables the study of most, if not all, nonstationary indicators. Assumption B could be also relaxed if we require $N$ to grow faster than $T$ in the asymptotics (to control for the contribution of series that are too persistent in factor projection), while under Assumption B the condition $T/N^2 \to 0$, also used by Pesaran (2006) for weakly dependent series, is needed in our analysis. There is no further requirement on the distribution of values of $\delta_i$ across individuals, apart from the restrictions on the maximum value in Assumption B. No restrictions on the lower value of $\varrho$ or $\vartheta_i$, nor on $\delta_{i0}$, are needed for estimation of $\boldsymbol{\beta}_{i0}$, but $\delta_{i0} > 0$ will be imposed for estimation of $\theta_{i0}$.

Let

$$\boldsymbol{\Upsilon}_{\beta_i} = \sigma_i^2 \boldsymbol{\Sigma}_{ie}^{-1}(0) \boldsymbol{\Sigma}_{i0} \boldsymbol{\Sigma}_{ie}^{-1}(0).$$

**Theorem 2.** *Under Assumptions* A *and* B, *and if* $TN^{-2} \to 0$ *as* $(N, T)_j \to \infty$, *for fixed i,*

$$T^{1/2}\left(\hat{\boldsymbol{\beta}}_i(\delta^*) - \boldsymbol{\beta}_{i0}\right) \to_d \mathcal{N}(0, \boldsymbol{\Upsilon}_{\beta_i}).$$

Note that if $\delta^* = \delta_{i0}$ and $\psi(L;\xi) = 1$, $\boldsymbol{\Upsilon}_{\beta_i} = \sigma_i^2 \boldsymbol{\Sigma}_{ie}^{-1}(0)$ because $\zeta_{i0}(j) = 1\{j = 0\}$, so this result shows that in this case, the estimate $\hat{\boldsymbol{\beta}}_i(\delta^*)$ effectively becomes an efficient GLS estimate and the asymptotic variance of $\hat{\boldsymbol{\beta}}_i(\delta^*)$ simplifies in the usual way, not depending on the dynamics of the exactly prewhitened error term. The rate of convergence is $\sqrt{T}$ for the range of allowed memory parameters (or if $\delta^*$ is large enough as described in Assumption B), irrespective of cointegration among idiosyncratic terms of the observables, as the GLS estimate is designed in terms of approximately independent and stationary regressor and error time series after factor removal. Consistent estimates of the asymptotic variance of $\hat{\boldsymbol{\beta}}_i(\delta^*)$ could be designed adapting the methods of Robinson and Hidalgo (1997) and Robinson (2005) in terms of projected observations after factor elimination and residual series.

### 2.1. Estimation of dynamic parameters

We now concentrate on individual long and short memory parameter estimation. Define

$$\hat{\theta}_i = \arg\min_{\theta \in \Theta_i} L^*_{i,T}(\theta),$$

minimizing the conditional sum of squares (CSS)

$$L^*_{i,T}(\theta) = \frac{1}{T} \boldsymbol{\varepsilon}_i(\theta)' \boldsymbol{\varepsilon}_i(\theta),$$

where

$$\boldsymbol{\varepsilon}_i(\theta) = \lambda\left(L; \delta - \delta^*, \xi\right)\left(\tilde{\mathbf{y}}_i(\delta^*) - \tilde{\mathbf{X}}_i(\delta^*)\hat{\boldsymbol{\beta}}_i(\delta^*)\right)$$

and the vectors of observations $\tilde{\mathbf{y}}_i(\delta^*) = \bar{\mathcal{W}}_T(\delta^*)\mathcal{Y}_i(\delta^*)$ and $\tilde{\mathbf{X}}_i(\delta^*) = \bar{\mathcal{W}}_T(\delta^*)\mathcal{X}_i(\delta^*)$ and the least squares coefficients $\hat{\boldsymbol{\beta}}_i(\delta^*)$ are obtained after projection of $\mathcal{Y}_i(\delta^*)$ and $\mathcal{X}_i(\delta^*)$ on their stacked cross-section averages for a given $\delta^*$.

The estimates are only implicitly defined and entail optimization over $\Theta_i = \mathcal{D}_i \times \Xi_i$, where $\Xi_i$ is a compact subset of $\mathbb{R}^p$ and $\mathcal{D}_i = [\underline{\delta}_i, \overline{\delta}_i]$, with $0 < \underline{\delta}_i < \overline{\delta}_i < 3/2$. Though the aim in this type of problems is to cover a wide range of values of $\delta_i \in \mathcal{D}_i$ in the asymptotics, c.f. Hualde and Robinson (2011) and Nielsen (2014), in our framework there are interactions with other model parameters that require to restrict the set $\mathcal{D}_i$ reflecting some a priori knowledge on the true value of $\delta_i$ or to introduce further assumptions on $N$ and $T$. Here, although the exclusion of $\delta_i = 0$ seems unnatural, negligibility of initial-condition terms that are of size $O_p(T^{-\delta_{min}})$ in the memory estimation step requires $\delta_{min} > 0$ which is guaranteed only when $\underline{\delta}_i > 0$. This condition is imposed in our framework because, unlike in most of the literature, we control for the effects of the initial conditions of the fractional filters; see the supplementary online material in Appendix A for further details.

In our analysis we impose the next assumption, which requires that $\underline{\delta}_i$ is not too small compared to the other memory parameters, implying that they cannot be very different from $\delta_{i0}$ which has to belong to the set $\mathcal{D}_i$.

**Assumption C1.** $\max\{\delta_{\max}, \vartheta_{\max}, \varrho\} - \underline{\delta}_i < 1/2$.

Here Assumption C1 basically imposes restrictions on the degree of heterogeneity and on the lower bound $\underline{\delta}_i$. Note that when $\delta_{i0} \in \mathcal{D}_i$ the conditions in Assumption C1 imply $\vartheta_i - \delta_{i0} < 1/2$ because $\vartheta_i \leq \vartheta_{\max}$ and $\underline{\delta}_i \leq \delta_{i0}$, and also imply $\varrho - \delta_{i0} < 1/2$. We also consider next a relaxed version of this condition that allows for a larger distance between the memory parameters and $\underline{\delta}_i$, and, in particular, permits lower values of $\delta_{i0}$ in the analysis when $\rho = 1$, i.e. the common factor is $I(1)$, in correspondence with the assumption $\delta_{i0} > \frac{1}{4}$ that will be imposed for our asymptotic normality result. However, it still requires a somewhat strong a priori knowledge on the whereabouts of $\delta_{i0}$ to fix $\underline{\delta}_i$.

**Assumption C2.** $\max\{\delta_{\max}, \vartheta_{\max}, \varrho\} - \underline{\delta}_i < 3/4$ and $\delta_i - \underline{\delta}_i < 1/2$.

Next, we establish the consistency and asymptotic normality of the dynamic parameter estimates, for which we further impose the following condition.

**Assumption D.** $\psi(L;\xi)$ is twice continuously differentiable for all $\xi \in \Xi$ with $\dot{\psi}_t(L;\xi) = (d/d\xi)\psi_t(L;\xi)$ where it is assumed that $\left|\dot{\psi}_t(L;\xi)\right| = O(\exp(-c(\xi)j))$.

Define

$$\begin{aligned}\chi(L;\xi) &= \frac{\partial}{\partial\theta}\log\lambda(L;\theta) = \left(\log\Delta, (\partial/\partial\xi')\log\psi(L;\xi)\right)' \\ &= \sum_{j=1}^{\infty}\chi_j(\xi)L^j,\end{aligned}$$

where $\chi_j(\xi) = \left(-1/j, \chi_{2j}(\xi)'\right)'$, and introduce the $(p+1) \times (p+1)$ matrix

$$\begin{aligned}B(\xi) &= \sum_{j=1}^{\infty}\chi_j(\xi)\chi_j'(\xi) \\ &= \begin{bmatrix} \pi^2/6 & -\sum_{j=1}^{\infty}\chi_{2j}'(\xi)/j \\ -\sum_{j=1}^{\infty}\chi_{2j}(\xi)/j & \sum_{j=1}^{\infty}\chi_{2j}(\xi)\chi_{2j}'(\xi)\end{bmatrix},\end{aligned}$$

and assume $B(\xi_{i0})$ is non-singular.

**Theorem 3.** *Under the assumptions of Theorem 2, Assumption C1 and D, $\theta_{i0} \in \Theta_i$ and $\delta_{i0} > 0$, for fixed i, $\hat{\theta}_i$ is consistent as $(N,T)_j \to \infty$, and if additionally $\delta_{i0} > \frac{1}{4}$ and $\theta_{i0} \in Int(\Theta_i)$, as $(N,T)_j \to \infty$,*

$$T^{1/2}\left(\hat{\theta}_i - \theta_{i0}\right) \to_d \mathcal{N}(0, B^{-1}(\xi_{i0})),$$

*where $\hat{\theta}_i$ is asymptotically uncorrelated with $\hat{\boldsymbol{\beta}}_i(\delta^*)$. The results remain true if we replace Assumption C1 by Assumption C2 and additionally $T/N \to 0$ as $(N,T)_j \to \infty$.*

Note that Theorem 2 guarantees the $\sqrt{T}$ consistency of $\hat{\boldsymbol{\beta}}_i(\delta^*)$, which might be stronger than needed for the consistency of $\hat{\theta}_i$, but simplifies the proof, while with Assumption C2 and the condition $TN^{-1} \to 0$ we allow for a wider range of values of $\rho$, $\delta_i$ and $\vartheta_i$, as is done in the analysis of the pooled estimate based on first differences of Robinson and Velasco (2015), where the condition $\delta_{i0} > \frac{1}{4}$ is also needed to control initial condition bias. The estimates of slope and dynamic parameters are asymptotically independent. The reason is that, after factor projection/estimation, the regressors are approximately strictly exogenous, since the idiosyncratic part of the regressors is independent of the idiosyncratic shocks at all leads and lags affected by the filter $\lambda(\cdot)$. However, note that the long and short memory parameter estimates are not asymptotically uncorrelated.

### 2.2. GLS estimation

In this section we show the efficiency of the feasible GLS slope estimate $\tilde{\boldsymbol{\beta}}_i(\hat{\theta}_i)$ obtained by plugging in an estimate, $\hat{\theta}_i$, of the vector $\theta_{i0}$, where $\hat{\theta}_i$ is $\sqrt{T}$ consistent for $\theta_{i0}$, with $\delta^* = \delta_{i0}$ satisfying the restrictions in Assumption B, though $\delta_{i0}$ is not known and only estimated jointly with the other dynamic parameters by $\hat{\theta}_i$. Note that this requires $\delta_{i0} \geq 1$ in a general set up where factors and the idiosyncratic component of regressors can have orders of integration arbitrarily close to 3/2. For that, define the following generalized prewhitened series,

$$\widehat{\mathcal{Y}}_j = \widehat{\mathcal{Y}}_j(\hat{\theta}_i) = \lambda_{t-1}(L;\hat{\theta}_i^{(-1)})\Delta \mathbf{y}_j$$

$$\widehat{\mathcal{X}}_j = \widehat{\mathcal{X}}_j(\hat{\theta}_i) = \lambda_{t-1}(L;\hat{\theta}_i^{(-1)})\Delta \mathbf{X}_j$$

for $j = 1,\ldots,N$, aiming to whiten also the short memory with $\hat{\theta}_i^{(-1)} = \left(\hat{\theta}_i - 1, \hat{\xi}_i'\right)'$, and their cross-section averages, $\hat{\mathbf{y}}(\hat{\theta}_i)$ and $\hat{\mathbf{X}}(\hat{\theta}_i)$, and the corresponding projection matrix $\widehat{\mathcal{W}}_T = \widehat{\mathcal{W}}_T(\hat{\theta}_i) = \mathbf{I}_T - \hat{\mathbf{H}}(\hat{\theta}_i)(\hat{\mathbf{H}}(\hat{\theta}_i)'\hat{\mathbf{H}}(\hat{\theta}_i))^{-}\hat{\mathbf{H}}(\hat{\theta}_i)'$ based on $\hat{\mathbf{H}}(\hat{\theta}_i) = N^{-1}\sum_{j=1}^N\left(\widehat{\mathcal{Y}}_j(\hat{\theta}_i)\ \widehat{\mathcal{X}}_j(\hat{\theta}_i)\right)$. Then the GLS estimate is

$$\tilde{\boldsymbol{\beta}}_i(\hat{\theta}_i) := \left(\widehat{\mathcal{X}}_i'\widehat{\mathcal{W}}_T\widehat{\mathcal{X}}_i\right)^{-1}\widehat{\mathcal{X}}_i'\widehat{\mathcal{W}}_T\widehat{\mathcal{Y}}_i$$

whose identification requires the following condition.

**Assumption A.6′**. The matrix $\widehat{\mathcal{X}}_i'\widehat{\mathcal{W}}_T\widehat{\mathcal{X}}_i$ is full rank.

Let

$$\bar{\boldsymbol{\Sigma}}_{ie} = \sum_{k=0}^{\infty}\bar{\Phi}_{ik}^{e}\boldsymbol{\Omega}_{ie}\bar{\Phi}_{ik}^{e\prime},$$

be the asymptotic variance matrix of the idiosyncratic component of the prewhitened regressors $\widehat{\mathcal{X}}_i^0 = \widehat{\mathcal{X}}_i(\theta_{i0})$ where the weights $\bar{\Phi}_{ik}^{e} = \sum_{j=0}^{k}\Phi_{ik-j}^{e}\lambda_j(\delta_{i0} - \vartheta_i, \xi_{i0})$ incorporate the prewhitening effect.

**Theorem 4.** *Assuming the conditions of Theorem 2 and Assumption A.6' to hold replacing $\delta^*$ with $\delta_{i0}$, and $\hat{\theta}_i - \theta_{i0} = O_p\left(T^{-1/2}\right)$, for fixed i,*

$$T^{1/2}\left(\tilde{\boldsymbol{\beta}}_i(\hat{\theta}_i) - \boldsymbol{\beta}_i\right) \to_d \mathcal{N}(0, \sigma_i^2\bar{\boldsymbol{\Sigma}}_{ie}^{-1}),$$

*where $\tilde{\boldsymbol{\beta}}_i(\hat{\theta}_i)$ is asymptotically uncorrelated with $\hat{\theta}_i$.*

Consistent estimation of $\sigma_i^2$ can be conducted directly from the sample variance of residuals $\boldsymbol{\varepsilon}_i(\hat{\theta}_i)$, while estimation of $\bar{\boldsymbol{\Sigma}}_{ie}$ would require the sample second moment matrix of the projected and prewhitened series regressors, i.e. $T^{-1}\widehat{\mathcal{X}}_i'\widehat{\mathcal{W}}_T\widehat{\mathcal{X}}_i$. Further iterations to estimate $\theta_i$ can also be envisaged using the efficient $\tilde{\boldsymbol{\beta}}_i(\hat{\theta}_i)$ instead of the preliminary $\hat{\boldsymbol{\beta}}_i(\delta^*)$, and then estimating $\beta_i$ again with the new $\hat{\theta}_i$.

### 2.3. Estimation of mean effects

Given the panel data structure, in many cases there is an interest in estimating the average effect across all cross section units. The simplest estimate capturing average effects is the common correlation mean group estimate that averages all individual coefficients, possibly with a common $\delta^*$,

$$\hat{\boldsymbol{\beta}}_{CCMG}(\delta^*) = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i(\delta^*).$$

Other possibilities such as the common correlation pooled estimate,

$$\hat{\boldsymbol{\beta}}_{CCP}(\delta^*) := \left(\sum_{i=1}^{N}\mathcal{X}_i'\bar{\mathcal{W}}_T\mathcal{X}_i\right)^{-1}\sum_{i=1}^{N}\mathcal{X}_i'\bar{\mathcal{W}}_T\mathcal{Y}_i,$$

can be more in the spirit of the joint estimation, possibly employing a common prewhitening parameter $\delta^*$. For the asymptotic analysis of the mean group estimate, we consider a simple linear random coefficients model

$$\beta_{i0} = \beta_0 + w_i, \quad w_i \sim iid\left(0, \Omega_w\right),$$

where $w_i$ is independent of all the other variables in the model. The asymptotic analysis of the pooled estimate requires further regularity conditions and is left for future research.

**Theorem 5.** *Under Assumptions A and B, and $\left(T^{-1}\mathcal{X}_i'\bar{\mathcal{W}}_T\mathcal{X}_i\right)^{-1}$ having finite second order moments for all $i = 1,\ldots,N$, as $(N,T)_j \to \infty$,*

$$\sqrt{N}\left(\hat{\boldsymbol{\beta}}_{CCMG}(\delta^*) - \boldsymbol{\beta}_0\right) \to_d \mathcal{N}(0, \Omega_w).$$

This theorem extends previous results in Pesaran (2006) and Kapetanios et al. (2011) for $I(0)$ and $I(1)$ variables under conditions similar to A.6 based on original data, where now the rate of convergence is $\sqrt{N}$, and no restrictions are required on the rate of growth of $N$ and $T$. Consistent estimates of the asymptotic variance can be proposed as in Pesaran (2006), since, asymptotically, variability only depends on the heterogeneity of the $\beta_{i0}$,

$$\hat{\Omega}_w = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_i(\delta^*) - \hat{\boldsymbol{\beta}}_{CCMG}(\delta^*)\right)\left(\hat{\boldsymbol{\beta}}_i(\delta^*) - \hat{\boldsymbol{\beta}}_{CCMG}(\delta^*)\right)'.$$

Similarly, the average effect can be estimated based on $\tilde{\boldsymbol{\beta}}_i(\hat{\theta}_i)$ as

$$\tilde{\boldsymbol{\beta}}_{CCMG}(\hat{\boldsymbol{\theta}}) = \frac{1}{N}\sum_{i=1}^{N}\tilde{\boldsymbol{\beta}}_i(\hat{\theta}_i), \quad \hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1,\ldots,\hat{\theta}_N\right),$$

which is also asymptotically normally distributed under Theorem 4 assumptions and boundedness of the second moments of $\left(T^{-1}\widehat{\mathcal{X}}_i'\widehat{\mathcal{W}}_T\widehat{\mathcal{X}}_i\right)^{-1}$, and its asymptotic variance–covariance matrix can be estimated by

$$\tilde{\Omega}_w = \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{\boldsymbol{\beta}}_i(\hat{\theta}_i) - \tilde{\boldsymbol{\beta}}_{CCMG}(\hat{\boldsymbol{\theta}})\right)\left(\tilde{\boldsymbol{\beta}}_i(\hat{\theta}_i) - \tilde{\boldsymbol{\beta}}_{CCMG}(\hat{\boldsymbol{\theta}})\right)'.$$

**Table 1**
Preliminary and joint estimation bias and RMSE's with $N = 10$ and $T = 50$ ($\delta^* = 1$).

| | | $\vartheta = 0.75$ | | | $\vartheta = 1$ | | | $\vartheta = 1.25$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\delta_0 = 0.5$ | $\delta_0 = 0.75$ | $\delta_0 = 1$ | $\delta_0 = 0.5$ | $\delta_0 = 0.75$ | $\delta_0 = 1$ | $\delta_0 = 0.5$ | $\delta_0 = 0.75$ | $\delta_0 = 1$ |
| $\varrho = 0.4$ : | | | | | | | | | | |
| Bias of $\hat{\beta}$ | $\hat{\beta}_{MG}(\delta^*)$ | 0.0016 | 0.0005 | 0.0023 | −0.0026 | −0.0058 | −0.0046 | −0.0086 | −0.0159 | −0.0179 |
| | $\hat{\beta}_{CC}(\delta^*)$ | 0.0012 | 0.0012 | 0.0012 | 0.0012 | 0.0012 | 0.0114 | 0.0005 | 0.0006 | 0.0009 |
| | $\hat{\beta}_{MG}(\hat{\delta})$ | 0.0007 | 0.0006 | 0.0006 | 0.0008 | 0.0008 | 0.0009 | 0.0005 | 0.0006 | 0.0010 |
| | $\hat{\beta}_{CC}(\hat{\delta})$ | 0.0149 | 0.0054 | 0.0014 | 0.0089 | 0.0044 | 0.0016 | 0.0028 | 0.0018 | 0.0011 |
| RMSE of $\hat{\beta}$ | $\hat{\beta}_{MG}(\delta^*)$ | 0.0621 | 0.0567 | 0.0529 | 0.0611 | 0.0573 | 0.0552 | 0.0538 | 0.0536 | 0.0555 |
| | $\hat{\beta}_{CC}(\delta^*)$ | 0.0621 | 0.0569 | 0.0531 | 0.0609 | 0.0571 | 0.0551 | 0.0518 | 0.0501 | 0.0518 |
| | $\hat{\beta}_{MG}(\hat{\delta})$ | 0.0621 | 0.0567 | 0.0529 | 0.0611 | 0.0570 | 0.0550 | 0.0531 | 0.0512 | 0.0525 |
| | $\hat{\beta}_{CC}(\hat{\delta})$ | 0.0589 | 0.0559 | 0.0531 | 0.0454 | 0.0520 | 0.0550 | 0.0293 | 0.0403 | 0.0517 |
| Bias of $\hat{\delta}$ | $\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$ | 0.0854 | 0.0218 | −0.0089 | 0.1133 | 0.0302 | −0.0083 | 0.1635 | 0.0488 | −0.0082 |
| | $\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$ | 0.0840 | 0.0211 | −0.0089 | 0.1100 | 0.0288 | −0.0083 | 0.1573 | 0.0462 | −0.0083 |
| RMSE of $\hat{\delta}$ | $\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$ | 0.0968 | 0.0458 | 0.0402 | 0.1245 | 0.0512 | 0.0399 | 0.1762 | 0.0673 | 0.0406 |
| | $\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$ | 0.0956 | 0.0456 | 0.0403 | 0.1217 | 0.0506 | 0.0401 | 0.1711 | 0.0660 | 0.0410 |
| $\varrho = 1$ : | | | | | | | | | | |
| Bias of $\hat{\beta}$ | $\hat{\beta}_{MG}(\delta^*)$ | −0.0029 | −0.0019 | 0.0017 | −0.0039 | −0.0052 | −0.0024 | −0.0070 | −0.0131 | −0.0140 |
| | $\hat{\beta}_{CC}(\delta^*)$ | 0.0006 | 0.0006 | 0.0008 | 0.0006 | 0.0007 | 0.0011 | 0.0001 | 0.0002 | 0.0007 |
| | $\hat{\beta}_{MG}(\hat{\delta})$ | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0002 | 0.0005 | 0.0001 | 0.0002 | 0.0006 |
| | $\hat{\beta}_{CC}(\hat{\delta})$ | 0.0436 | 0.0145 | 0.0012 | 0.0327 | 0.0127 | 0.0015 | 0.0146 | 0.0067 | 0.0012 |
| RMSE of $\hat{\beta}$ | $\hat{\beta}_{MG}(\delta^*)$ | 0.0624 | 0.0573 | 0.0537 | 0.0617 | 0.0580 | 0.0559 | 0.0545 | 0.0539 | 0.0555 |
| | $\hat{\beta}_{CC}(\delta^*)$ | 0.0626 | 0.0577 | 0.0541 | 0.0618 | 0.0581 | 0.0563 | 0.0533 | 0.0517 | 0.0534 |
| | $\hat{\beta}_{MG}(\hat{\delta})$ | 0.0624 | 0.0573 | 0.0537 | 0.0616 | 0.0577 | 0.0559 | 0.0540 | 0.0523 | 0.0537 |
| | $\hat{\beta}_{CC}(\hat{\delta})$ | 0.1033 | 0.0678 | 0.0539 | 0.0873 | 0.0648 | 0.0562 | 0.0577 | 0.0516 | 0.0533 |
| Bias of $\hat{\delta}$ | $\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$ | 0.1735 | 0.0609 | 0.0030 | 0.1870 | 0.0661 | 0.0033 | 0.2196 | 0.0816 | 0.0049 |
| | $\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$ | 0.1724 | 0.0600 | 0.0031 | 0.1868 | 0.0651 | 0.0033 | 0.2179 | 0.0800 | 0.0049 |
| RMSE of $\hat{\delta}$ | $\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$ | 0.1903 | 0.0821 | 0.0427 | 0.2017 | 0.0862 | 0.0430 | 0.2327 | 0.1003 | 0.0451 |
| | $\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$ | 0.1891 | 0.0816 | 0.0429 | 0.2010 | 0.0855 | 0.0433 | 0.2309 | 0.0991 | 0.0454 |

Note: This table reports the estimation results for the memory parameter as well as the mean group and common correlated slope parameters based on $\delta^*$ and $\hat{\delta}$. See Theorems 2–5 for the corresponding theoretical results.

Note that in Theorem 5, Assumption A.4 can be dropped as Pesaran (2006) discusses in length. The mean-group estimate is still consistent and asymptotically normally distributed in that case although its asymptotic variance would be different, which could nevertheless be estimated nonparametrically based on the variability of $\beta_i$. Under the rank condition in Assumption A.4, on the other hand, it is no longer required that the factor loadings follow a random coefficients model, and it suffices that they are bounded.

## 3. Monte Carlo simulations

In this section we carry out a Monte Carlo experiment to study the small-sample performance of the slope and memory estimates in the simplest case in which there is no short memory dynamics, $\xi = 0$, and persistence depends only on the value of $\delta_0$. We draw the idiosyncratic shocks $\varepsilon_{i,t}$ as standard normals and the factor loadings $\gamma_i$ from $U(-0.5, 1)$ not to restrict their sign. We then generate serially correlated common factors $f_t$ based on the *iid* shocks drawn as standard normals and then fractionally integrated to the order $\varrho$. The individual effects $\alpha_i$ are left unspecified since they are removed via first differencing in the estimation, and projections are based on the first-differenced data. We focus on different cross-section and time-series sizes, $N$ and $T$, as well as different values of $\delta_0$. Simulations are based on 1000 replications.

We conduct a finite-sample study to check the accuracy of both slope and fractional parameter estimates. The idiosyncratic component of covariates follows a pure fractional process of memory $\vartheta$. We investigate the performance for $(N, T) = (10, 50)$ and $(N, T) = (20, 100)$ for the parameter values $\delta_0 = 0.5, 0.75, 1$; $\vartheta = 0.75, 1, 1.25$, and $\varrho = 0.4, 1$, covering both cointegration (e.g. $\vartheta = 1.25$ and $\delta = 1$) and non-cointegration cases (e.g. $\vartheta = 1$ and $\delta = 1$). For projection of estimated factors based on prewhitened cross section averages, we take $\delta^* = 1$.

Tables 1 and 2 present biases and RMSE's for both slope and fractional parameter estimates for $(N, T) = (10, 50), (20, 100)$, respectively. Biases of both common correlation pooled (CCP) and mean group (CCMG) estimates are very reasonable with biases of pooled estimates generally dominating those of MG estimates, particularly when $\varrho = 1$. Biases of slope estimates become negative with their magnitudes increasing with $NT$ for the two smallest values of $\vartheta$. The pooled estimate of the fractional parameter suffers from large biases when $\delta_0$ is small relative to $\vartheta$ or $\varrho$ due to the idiosyncratic shocks in the regression equation being dominated by other sources of persistence. As expected, biases in fractional parameter estimates decrease with $\delta_0$ in all cases.

In terms of performance, slope estimates behave quite well both in cointegration and non-cointegration cases implying that cointegration is not necessary for the estimation of slope in practice. However, for several cases, the RMSE of fractional parameter estimates is rather large, which can be explained by persistence distortions from the common factor and covariate shocks. Nevertheless, performance of both slope and fractional parameter estimates is clearly improving with $\delta_0$ when $\vartheta = 0.75, 1$ and in all cases with $NT$. Efficiency gains of GLS type of estimates using $\hat{\delta}$ are very small, if any, for the MG estimate for all values of $\delta_0$, but for $\delta_0 < \delta^* = 1$ the behavior of the CCP estimate can deteriorate substantially, so overdifferencing in the prewhitening step seems a safe recommendation in practice.

## 4. A fractional panel analysis of realized volatilities

The capital asset pricing model (CAPM) and its variations have long been used in finance to determine a theoretically appropriate rate of return in a diversified portfolio, where estimating *beta* is essential as it measures the sensitivity of expected excess stock returns to expected excess market returns. While CAPM and other such models may prove useful in an $I(0)$ environment, they fail to

**Table 2**
Preliminary and joint estimation bias and RMSE's with $N = 20$ and $T = 100$ ($\delta^* = 1$).

| | | $\vartheta = 0.75$ | | | $\vartheta = 1$ | | | $\vartheta = 1.25$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\delta_0 = 0.5$ | $\delta_0 = 0.75$ | $\delta_0 = 1$ | $\delta_0 = 0.5$ | $\delta_0 = 0.75$ | $\delta_0 = 1$ | $\delta_0 = 0.5$ | $\delta_0 = 0.75$ | $\delta_0 = 1$ |
| $\varrho = 0.4:$ | | | | | | | | | | |
| Bias of $\hat{\beta}$ | $\hat{\beta}_{MG}(\delta^*)$ | −0.0022 | −0.0013 | −0.0009 | 0.0004 | 0.0015 | 0.0016 | 0.0058 | 0.0074 | 0.0080 |
| | $\hat{\beta}_{CC}(\delta^*)$ | −0.0011 | −0.0013 | −0.0014 | −0.0011 | −0.0014 | −0.0017 | −0.0006 | −0.0011 | −0.0017 |
| | $\hat{\beta}_{MG}(\hat{\delta})$ | −0.0011 | −0.0013 | −0.0014 | −0.0010 | −0.0013 | −0.0017 | −0.0006 | −0.0010 | −0.0016 |
| | $\hat{\beta}_{CC}(\hat{\delta})$ | 0.0136 | 0.0026 | −0.0013 | 0.0076 | 0.0018 | −0.0017 | 0.0022 | 0.0005 | −0.0016 |
| RMSE of $\hat{\beta}$ | $\hat{\beta}_{MG}(\delta^*)$ | 0.0295 | 0.0270 | 0.0254 | 0.0290 | 0.0271 | 0.0265 | 0.0256 | 0.0251 | 0.0262 |
| | $\hat{\beta}_{CC}(\delta^*)$ | 0.0299 | 0.0274 | 0.0258 | 0.0296 | 0.0276 | 0.0269 | 0.0251 | 0.0241 | 0.0252 |
| | $\hat{\beta}_{MG}(\hat{\delta})$ | 0.0294 | 0.0270 | 0.0254 | 0.0290 | 0.0271 | 0.0265 | 0.0250 | 0.0240 | 0.0250 |
| | $\hat{\beta}_{CC}(\hat{\delta})$ | 0.0341 | 0.0279 | 0.0258 | 0.0239 | 0.0258 | 0.0269 | 0.0131 | 0.0189 | 0.0252 |
| Bias of $\hat{\delta}$ | $\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$ | 0.0681 | 0.0174 | −0.0028 | 0.0984 | 0.0257 | −0.0012 | 0.1640 | 0.0490 | 0.0019 |
| | $\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$ | 0.0679 | 0.0173 | −0.0028 | 0.0975 | 0.0253 | −0.0012 | 0.1616 | 0.0482 | 0.0019 |
| RMSE of $\hat{\delta}$ | $\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$ | 0.0723 | 0.0259 | 0.0189 | 0.1046 | 0.0329 | 0.0187 | 0.1739 | 0.0573 | 0.0195 |
| | $\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$ | 0.0721 | 0.0259 | 0.0189 | 0.1038 | 0.0327 | 0.0187 | 0.1720 | 0.0568 | 0.0195 |
| $\varrho = 1:$ | | | | | | | | | | |
| Bias of $\hat{\beta}$ | $\hat{\beta}_{MG}(\delta^*)$ | −0.0031 | −0.0026 | −0.0027 | 0.0001 | 0.0008 | 0.0003 | 0.0068 | 0.0082 | 0.0082 |
| | $\hat{\beta}_{CC}(\delta^*)$ | −0.0013 | −0.0015 | −0.0015 | −0.0013 | −0.0016 | −0.0019 | −0.0009 | −0.0013 | −0.0018 |
| | $\hat{\beta}_{MG}(\hat{\delta})$ | −0.0013 | −0.0015 | −0.0016 | −0.0012 | −0.0015 | −0.0018 | −0.0008 | −0.0012 | −0.0018 |
| | $\hat{\beta}_{CC}(\hat{\delta})$ | 0.0588 | 0.0155 | −0.0015 | 0.0423 | 0.0130 | −0.0018 | 0.0159 | 0.0062 | −0.0017 |
| RMSE of $\hat{\beta}$ | $\hat{\beta}_{MG}(\delta^*)$ | 0.0297 | 0.0273 | 0.0258 | 0.0293 | 0.0274 | 0.0267 | 0.0263 | 0.0258 | 0.0267 |
| | $\hat{\beta}_{CC}(\delta^*)$ | 0.0302 | 0.0277 | 0.0261 | 0.0300 | 0.0280 | 0.0273 | 0.0258 | 0.0248 | 0.0259 |
| | $\hat{\beta}_{MG}(\hat{\delta})$ | 0.0296 | 0.0272 | 0.0257 | 0.0293 | 0.0274 | 0.0268 | 0.0255 | 0.0245 | 0.0255 |
| | $\hat{\beta}_{CC}(\hat{\delta})$ | 0.0927 | 0.0403 | 0.0260 | 0.0713 | 0.0371 | 0.0272 | 0.0362 | 0.0264 | 0.0258 |
| Bias of $\hat{\delta}$ | $\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$ | 0.1383 | 0.0406 | 0.0017 | 0.1545 | 0.0468 | 0.0032 | 0.2019 | 0.0680 | 0.0074 |
| | $\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$ | 0.1390 | 0.0404 | 0.0017 | 0.1570 | 0.0466 | 0.0032 | 0.2028 | 0.0676 | 0.0074 |
| RMSE of $\hat{\delta}$ | $\hat{\delta}(\hat{\beta}_{CC}(\delta^*))$ | 0.1479 | 0.0494 | 0.0194 | 0.1628 | 0.0548 | 0.0198 | 0.2103 | 0.0765 | 0.0224 |
| | $\hat{\delta}(\hat{\beta}_{CC}(\hat{\delta}))$ | 0.1482 | 0.0491 | 0.0195 | 0.1646 | 0.0546 | 0.0198 | 0.2107 | 0.0761 | 0.0224 |

Note: This table reports the estimation results for the memory parameter as well as the mean group and common correlated slope parameters based on $\delta^*$ and $\hat{\delta}$. See Theorems 2–5 for the corresponding theoretical results.

provide valid inference for variables that exhibit fractional long-range dependence such as volatility.

In this application, we assess the sensitivity of industry realized volatilities to a market realized volatility measure. In particular, we estimate the betas for volatility under our general setup, which allows for possible cointegrating relationships. Such long-run relationships may have direct policy and investment implications since they provide information as to which industries are susceptible to a potential market risk upheaval. Bearing in mind an economy as a portfolio of industries, we use our general model to assess the presence and effects of systematic risk in an economy.

In order to calculate monthly realized volatility measures, we use daily average-value-weighted returns data spanning the time period 2000–2011 ($T = 144$ months) from Kenneth French's Data Library for 30 industries in the U.S. economy. As for the composite market returns, we use a weighted average of daily returns of NYSE, NASDAQ and AMEX since the companies considered in industry returns trade in one of these markets. Using the composite index returns of NYSE, NASDAQ and AMEX, i.e. $r_{m,t}$, we calculate

$$RVM_t = \left(\sum_{s \in t}^{N_t} r_{m,s}^2\right)^{1/2}, \quad t = 1, 2, \ldots, T,$$

where $N_t$ is the number of trading (typically 22) days in a month. Next, for each industry, we calculate

$$RVI_{i,t} = \left(\sum_{s \in t}^{N_t} e_{i,s}^2\right)^{1/2}, \quad t = 1, 2, \ldots, T,$$

where $e_{i,s} = r_{i,s} - r_{m,s}$, cf. Chauvet et al. (2012). Along this line, while jump-robust measures such as bipower variation could also be used, our sole purpose with this empirical application is to show that our general model is suited to address the empirical problem described herein.

Fig. 1 shows the behavior of monthly industry realized volatilities and justifies a heterogeneous approach. Fig. 2 shows the realized volatility in the composite average of NYSE, NASDAQ and AMEX, where especially closer to the spike there is a trending behavior also shared by some of the industries as seen in Fig. 1.

Observing that the volatility of volatility is time-varying, we scale each industry as well as the market realized volatility by their corresponding standard deviations. Then we estimate

$$RVI_{i,t} = \alpha_i + \beta_{i0}^0 RVM_t + \beta_{i0} X_{i,t} + \gamma_i' f_t + \Delta_{t+1}^{-\delta_i} v_{i,t},$$

where $RVM_t$, the $I(\vartheta)$ market realized volatility, is the observable common factor that is treated as a covariate; each $X_{i,t}$ is the average effect of $I(0)$ industry-specific factors: book-to-market ratio and market capitalization, which are also covariates; $f_t$ are $I(\varrho)$ unobservable common factors that are projected out as described in earlier sections so that possible cointegrating relationships can be disclosed between $RVI_{i,t}$ and $RVM_t$.

We obtain fractional integration degrees of market and industry realized volatilities resorting to local Whittle estimation, Robinson (1995), with bandwidth choices of $m = T^{0.6}, T^{0.7}$ corresponding to $m = 20, 32$, respectively, and refrain from adding more Fourier frequencies to avoid higher-frequency contamination. Table 3 collectively presents the local Whittle estimates of fractional integration values of the 30 U.S. industry realized volatilities as well as those of the composite market. For both bandwidth choices, the industry realized volatilities display heterogeneity lying above the nonstationarity bound. The market realized volatility is also nonstationary being integrated of an order around 0.6. The unobserved common factor has integration orders of $\varrho =$

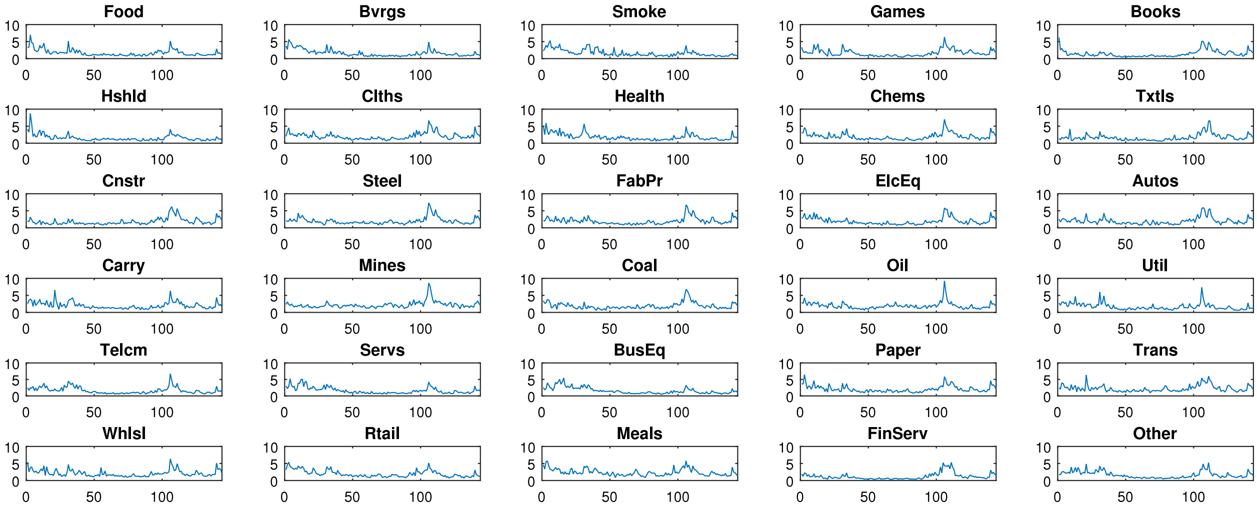


**Fig. 1.** Monthly realized volatilities across industries.

**Table 3**
Estimated integration orders of industry realized volatilities.

| $m = 20$ : | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Food | Bvrgs | Tobac | Games | Books | Hshld | Clths | Hlth | Chems | Txtls | **Market** |
| 0.51 | 0.77 | 0.71 | 0.75 | 0.84 | 0.51 | 0.70 | 0.72 | 0.68 | 0.69 | 0.59 |
| Cnstr | Steel | FabPr | ElcEq | Autos | Carry | Mines | Coal | Oil | Util | |
| 0.73 | 0.71 | 0.73 | 0.86 | 0.74 | 0.70 | 0.47 | 0.71 | 0.56 | 0.52 | |
| Telcm | Servs | BusEq | Paper | Trans | Whlsl | Rtail | Meals | Finan | Other | |
| 0.83 | 0.66 | 0.85 | 0.78 | 0.61 | 0.52 | 0.67 | 0.56 | 0.98 | 0.77 | |
| $m = 32$ : | | | | | | | | | | |
| Food | Bvrgs | Tobac | Games | Books | Hshld | Clths | Hlth | Chems | Txtls | **Market** |
| 0.66 | 0.78 | 0.63 | 0.57 | 0.63 | 0.46 | 0.60 | 0.71 | 0.67 | 0.59 | 0.64 |
| Cnstr | Steel | FabPr | ElcEq | Autos | Carry | Mines | Coal | Oil | Util | |
| 0.74 | 0.72 | 0.64 | 0.69 | 0.56 | 0.55 | 0.54 | 0.63 | 0.58 | 0.58 | |
| Telcm | Servs | BusEq | Paper | Trans | Whlsl | Rtail | Meals | Finan | Other | |
| 0.79 | 0.75 | 0.78 | 0.60 | 0.57 | 0.62 | 0.77 | 0.57 | 0.90 | 0.78 | |

Note: This table reports the local Whittle estimation results of the individual integration orders of industry and market realized volatilities with bandwidth choices of $m = 20, 32$. Estimates are rounded to two digits after zero. Standard errors of the estimates are 0.112 and 0.088 respectively for $m = 20, 32$.

**Table 4**
Residual integration order estimates ($\hat{\delta}_i$) of industry realized volatilities.

| Food | Bvrgs | Tobac | Games | Books | Hshld | Clths | Hlth | Chems | Txtls |
|---|---|---|---|---|---|---|---|---|---|
| 0.50 | 0.54 | 0.49 | 0.48 | 0.59 | 0.54 | 0.30 | 0.50 | 0.42 | 0.40 |
| Cnstr | Steel | FabPr | ElcEq | Autos | Carry | Mines | Coal | Oil | Util |
| 0.48 | 0.50 | 0.30 | 0.50 | 0.30 | 0.29 | 0.45 | 0.48 | 0.50 | 0.37 |
| Telcm | Servs | BusEq | Paper | Trans | Whlsl | Rtail | Meals | Finan | Other |
| 0.51 | 0.58 | 0.65 | 0.43 | 0.42 | 0.28 | 0.65 | 0.54 | 0.53 | 0.43 |

Note: This table reports the estimation results of the integration order of individual industry realized volatility residuals. Estimations are performed based on our general model where the projections are carried out with $\delta^* = 1$. Values are rounded to two digits after zero. Standard error of these estimates is 0.065.

0.71, 0.66 for $m = 20, 32$, respectively, which we estimate based on the cross-section averages of the industry realized volatilities.

We use our general model to jointly estimate the fractional order of residuals ($\delta_i$) and slope coefficients ($\beta_{i0}^0$ and $\beta_{i0}$) based on the projections of first-differenced data ($\delta^* = 1$) in order to be able to confirm and identify cointegrating relationships. Fama–French factors are taken to be $I(0)$ in asset-pricing literature, rendering cointegration possible only between the market and industry realized volatilities. Table 4 presents the fractional order of residuals, from which the cointegrating relationships are confirmed based on the results presented in Table 3.

The main criterion for cointegration in this setup is $\delta_i < \vartheta_i$ since the equality of realized volatility integration orders between industries and the market cannot be rejected in all but very few cases. Based on these two requirements together, cointegrating relationships are confirmed between the market realized volatility and the realized volatilities of all industries but Financial Services, Business Equipment and Telecommunications for $m = 20$. With the bandwidth of $m = 32$, more pronounced cointegrating relationships with the market realized volatility are indicated for the realized volatilities of all industries except Financial Services. Estimates of the cointegrating parameters and their robust standard errors calculated from the asymptotic variance in Theorem 2 are reported in Table 5, from which it is obvious that the market realized volatility has a positive and significant effect on all industry realized volatilities with heterogeneous magnitudes while the average effect of industry characteristics (captured by Fama–French factors) displays differences across industries. Although for several industries slope parameters are estimated under non-cointegrating relationships, the finite-sample study in the previous section indicates that these estimates are still reliable.

**Table 5**
Estimated slope parameters across industry realized volatilities.

| | Food | Bvrgs | Tobac | Games | Books | Hshld | Clths | Hlth |
|---|---|---|---|---|---|---|---|---|
| $\hat{\beta}_i^0$ | 0.5422 | 0.4002 | 0.3376 | 0.6896 | 0.6503 | 0.2707 | 0.7446 | 0.4289 |
| | (0.1097) | (0.1379) | (0.1452) | (0.0762) | (0.0769) | (0.1234) | (0.0607) | (0.1199) |
| $\hat{\beta}_i$ | 1.8145 | 1.4060 | −0.1814 | 0.1361 | 0.4119 | −0.2088 | 2.4219 | −0.6377 |
| | (0.0856) | (0.1006) | (0.1328) | (0.0559) | (0.1144) | (0.0864) | (0.0602) | (0.0830) |
| | Cnstr | Steel | FabPr | ElcEq | Autos | Carry | Mines | Coal |
| $\hat{\beta}_i^0$ | 0.7346 | 0.8571 | 0.9094 | 0.6970 | 0.8332 | 0.6176 | 0.8373 | 0.7691 |
| | (0.0821) | (0.0633) | (0.0413) | (0.0758) | (0.0523) | (0.0814) | (0.0854) | (0.0807) |
| $\hat{\beta}_i$ | −0.4109 | 0.1789 | −0.4298 | −0.3442 | −0.3635 | 1.7414 | −0.5087 | 0.3626 |
| | (0.1266) | (0.0782) | (0.0537) | (0.0768) | (0.0765) | (0.0772) | (0.1335) | (0.1219) |
| | Telcm | Servs | BusEq | Paper | Trans | Whlsl | Rtail | Meals |
| $\hat{\beta}_i^0$ | 0.7190 | 0.6178 | 0.5250 | 0.6223 | 0.6183 | 0.8722 | 0.4078 | 0.5382 |
| | (0.0961) | (0.1271) | (0.1530) | (0.0768) | (0.0751) | (0.0603) | (0.1308) | (0.1020) |
| $\hat{\beta}_i$ | 0.1399 | −0.3669 | 0.0311 | −1.0433 | −0.1778 | −2.4097 | 2.6804 | −0.6838 |
| | (0.0628) | (0.1329) | (0.1718) | (0.0686) | (0.1065) | (0.1122) | (0.0832) | (0.0820) |
| | | Chems | Txtls | Oil | Util | Finan | Other | |
| $\hat{\beta}_i^0$ | | 0.7898 | 0.4888 | 0.7927 | 0.6498 | 0.5316 | 0.1067 | |
| | | (0.0516) | (0.0981) | (0.0852) | (0.0925) | (0.0986) | (0.0632) | |
| $\hat{\beta}_i$ | | −0.0546 | −0.1731 | −0.1238 | −0.4930 | −0.8456 | −0.1933 | |
| | | (0.0419) | (0.1665) | (0.0982) | (0.0828) | (0.1838) | (0.0881) | |

Note: This table reports the estimation results of the individual slope parameters across industry realized volatilities, where $\hat{\beta}_i^0$ is the coefficient of market realized volatility, and $\hat{\beta}_i$ is the coefficient of the average effect of Fama–French factors. Estimations are performed based on our general model where the projections are carried out with $\delta^* = 1$. Robust standard errors are reported in parentheses.

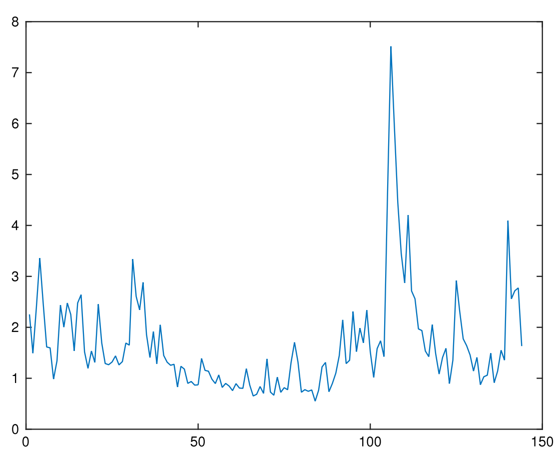


**Fig. 2.** Monthly realized volatility in the composite market.

This empirical study reveals that our general model can be used to assess the relationship between market and industry realized volatilities. In fact, other types of such nonstationarity assessment can be performed using our general model. Further studies may focus on estimating cointegrating vectors in-between industries to exactly identify the industries that could be safe to invest in during crises periods as well as to be able to foresee a potential crisis through the real sector.

## 5. Final comments

We have considered a large $N, T$ panel data model with fixed effects and cross-section dependence where the idiosyncratic shocks and common factors are allowed to exhibit long-range dependence. Our methodology for memory estimation consists in CSS estimation on the first differences of defactored variables, where projections are carried out on the sample means of differenced data, and slope estimation is carried out in the least-squares sense. While our methodology offers a general treatment for stationary and nonstationary indicators and works well in practice as indicated by Monte Carlo experiments, it can nevertheless be extended in the following directions: (a) Different estimation techniques, such as fixed effects and GMM, can be used under our setup as in Robinson and Velasco (2015); (b) The idiosyncratic shocks may be allowed to feature spatial dependence providing further insights in empirical analyses; (c) The independence assumption between the idiosyncratic shocks in the general model can be relaxed to allow for nonfactor endogeneity thereby leading to a cointegrated system analysis in the classical sense as in Ergemen (2015) who considers a less flexible modelization due to the lack of allowance of multiple covariates; (d) Panel unit-root and related hypothesis testing can be readily performed using our methodology, but it could also be interesting to develop tests that can detect breaks in the general model parameters; (e) Homogeneity tests on the slope parameters could be developed by comparing our mean group estimates with pooled estimates derived from a homogeneous version of our model.


## Acknowledgments

The authors are grateful to a co-editor, an associate editor and three anonymous referees whose helpful suggestions and constructive comments have led to several improvements in the clarity of exposition. The authors also would like to thank Manuel Arellano, Jesús Gonzalo, Javier Hualde, Bent Nielsen, Peter M. Robinson, Abderrahim Taamouti, the participants in the 67th Econometric Society European Meeting, CREATES Symposium on Long Memory 2013, Robust Econometric Methods for Modeling Economic and Financial Variables Conference 2012, UC3M Seminars and $\text{III}_t$ Workshop in Time Series Econometrics for helpful comments and discussions. Financial support from the Ministerio de Economía y Competitividad (Spain), grants ECO2012-31748, ECO2014-57007p and MDM 2014-0431, and Comunidad de Madrid, MadEco-CM (S2015/HUM-3444) is gratefully acknowledged. The authors also acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation (DNRF78).

## References


Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77(4), 1229–1279.

Bai, J., 2010. Panel unit root tests with cross-section dependence: A further investigation. Econometric Theory 26, 1088–1114.

Bai, J., Ng, S., 2004. A PANIC attack on unit roots and cointegration. Econometrica 72(4), 1127–1177.

Baillie, R., 1996. Long memory processes and fractional integration in econometrics. J. Econometrics 73, 5–59.

Bollerslev, T., Osterrieder, D., Sizova, N., Tauchen, G., 2013. Risk and return: Long-run relationships, fractional cointegration, and return predictability. J. Financ. Econ. 108(2), 409–424.

Chauvet, M., Senyuz, Z., Yoldas, E., 2012. What does realized volatility tell us about macroeconomic fluctuations? In: Finance and Economics Discussion Series. Board of Governors of the Federal Reserve System (U.S.).

Ergemen, Y.E., 2015. System Estimation of Panel Data Models under Long-Range Dependence. Preprint Aarhus University.

Gil-Alaña, L., Robinson, P., 1997. Testing of unit root and other nonstationary hypotheses in macroeconomic time series. J. Econometrics 80 (2), 241–268.

Granger, C., 1980. Long memory relationships and the aggregation of dynamic models. J. Econometrics 14, 227–238.

Hassler, U., Demetrescu, M., Tarcolea, A.I., 2011. Asymptotic normal tests for integration in panels with cross-dependent units. AStA Adv. Stat. Anal. 95, 187–204.

Hualde, J., Robinson, P.M., 2011. Gaussian pseudo-maximum likelihood estimation of fractional time series models. Ann. Statist. 39 (6), 3152–3181.

Kapetanios, G., Pesaran, M.H., Yamagata, T., 2011. Panels with non-stationary multifactor error structures. J. Econometrics 160 (2), 326–348.

Moon, H.R., Perron, B., 2004. Testing for a unit root in panels with dynamic factors. J. Econometrics 122 (1), 81–126.

Moon, H.R., Weidner, M., 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica 83 (4), 1543–1579.

Nielsen, M. Ø., 2014. Asymptotics for the conditional-sum-of-squares estimator in multivariate fractional time series models. J. Time Series Anal. http://dx.doi.org/10.1111/jtsa.12100.

Pesaran, H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74 (4), 967–1012.

Robinson, P.M., 1978. Statistical inference for a random coefficient autoregressive model. Scand. J. Stat. 5, 163–168.

Robinson, P.M., 1995. Gaussian semiparametric estimation of long range dependence. Ann. Statist. 23 (5), 1630–1661.

Robinson, P.M., 2005. Robust covariance matrix estimation : 'HAC' estimates with long memory/antipersistence correction. Econometric Theory 21 (1), 171–180.

Robinson, P.M., Hidalgo, J., 1997. Time series regression with long-range dependence. Ann. Stat. 25 (1), 77–104.

Robinson, P.M., Hualde, J., 2003. Cointegration in fractional systems with unknown integration orders. Econometrica 71 (6), 1727–1766.

Robinson, P.M., Velasco, C., 2015. Efficient estimation of fractionally integrated panel data models with fixed effects. J. Econometrics 185, 435–452.